NUCLEAR THEORY, Vol. 39 (2022)
eds. M. Gaidarov, N. Minkov, Heron Press, Sofia

Combining LASSO Regularization with Information Criteria: a Study of the $K^+\Sigma^-$ Photo-production with an Isobar Model

D. Petrellis[1], P. Bydžovský[1], A. Cieplý[1], D. Skoupil[1], N. Zachariou[2]

[1]Nuclear Physics Institute, Czech Academy of Sciences, Řež, Czech Republic
[2]University of York, York YO10 5DD, United Kingdom

Abstract. We investigate the $K^+\Sigma^-$ photoproduction by fitting an isobar model to experimental data. However, the large number of parameters involved (corresponding to prospective resonances), make the fitting procedure problematic. We propose using regularization, a machine learning method that is customarily used to prevent overfitting the data. Combined with model selection criteria, this process effectively leads to an optimal selection of the resonances involved in the reaction mechanism.

1 Introduction

The focus of the current study is the reaction $\gamma n \to K^+\Sigma^-$. Such reactions play an important role in our understanding of the baryon spectrum; however, due to a plethora of nucleon resonances in the energy region of interest (W $\approx$ $1.6 - 2.6$ GeV) one is faced with a large number of fitting parameters and, most importantly, a huge number of possible combinations of resonances to consider - hence the need for *model selection* criteria, where by *model* we refer to an optimal subset out of a group of candidate resonances. In section 3 we present a method that has been used recently in similar contexts [1–3] and that comprises some well-established statistical learning techniques [4]. The results presented here have been reported in [5], while the data used are from [6] and [7].

2 The Isobar Approach

Due to their phenomenological nature, isobar models are based on several approximations [8, 9]. Their most important assumptions are that hadrons are the fundamental degrees of freedom and their interactions are given in terms of effective Lagrangians, while reaction amplitudes are constructed as sums of lowest-order Feynman diagrams. Higher-order processes are implicitly taken into account in *effective* couplings, which are obtained by fits to experimental data. Three types of diagrams are possible at this level of approximation, depending on the nature of the exchanged particle (nucleon, kaon or hyperon), each corresponding to the **s**, **t** and **u** Mandelstam variables. These diagrams are further distinguished according to whether the exchanged particle is in its ground, or in a resonant state (N^*, K^*, Y^*) leading to their characterization as Born or non-Born diagrams, respectively. Considering these six types of interactions, the exchanges of N^* or Δ^* (the **s**, non-Born diagram) produce resonant structures in the cross-section, while the rest of the diagrams contribute only to the "background".

3 Regularized Least-Squares Fitting and Information Criteria

One of the central issues in statistical modeling is optimal model complexity. Models that are too complex (*e.g.* with more parameters than necessary) tend to overfit the data, as opposed to simple models that underfit the data and lead to large error values. Overfitting is the result of fitting the noise in the sample, leading to a model of very low predictive power. A common way to deal with this problem is through the technique known as regularization, where the error function is redefined as

$$\chi_T^2 = \chi^2 + P(\lambda), \tag{1}$$

with

$$\chi^2 = \sum_{i=1}^{N} \frac{[d_i - p_i(w_1, \ldots, w_k)]^2}{(\sigma_{d_i}^{stat})^2}, \tag{2}$$

the ordinary χ^2 error function, while

$$P(\lambda) = \lambda^4 \sum_{j=1}^{k} |w_j|^q, \tag{3}$$

a penalty term containing the k parameters $\{w_1, w_2 \ldots w_k\}$, with respect to which the χ^2 function is minimized.

The presence of the $P(\lambda)$ term in the regularized error function (χ_T^2) imposes a constraint on the parameter values and effectively prevents overfitting. The magnitude of the λ regularization parameter determines the strength of the constraint, while the power q determines its type. The fact that λ is raised to the power of 4 allows us to focus more on the region of small λ's. In the case of $q = 1$, which is known in the literature as Least Absolute Shrinkage and Selection Operator (LASSO) regularization [4], some of the parameters are shrunk to zero. In this case, λ determines how sparse the resulting model is.

At this point, one can apply the information criteria (IC), namely the Akaike (AIC) [10] and Bayesian (BIC) [11], that help to determine the optimal value of λ, as the one that minimizes the corresponding expressions: AIC $= 2k + \chi_T^2$ and BIC $= k\log(N) + \chi_T^2$, where k is the number of parameters of the model and N is the size of the data set. The result of this process can be seen in Figure 1. Apparently, all criteria give similar results that differ only by a scale factor.

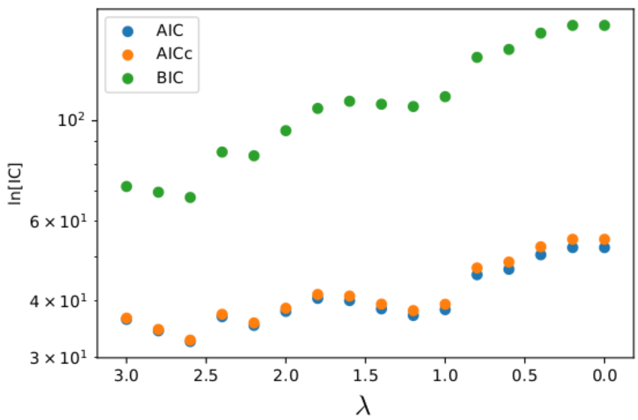

Figure 1. The Akaike (blue), a modified version known as *corrected Akaike* (orange), and Bayesian (green) information criteria for various values of the regularization parameter λ. Please, note the logarithmic scale of the vertical axis. (Taken from [5]).

4 Numerical Results and Discussion

In the present work, two fits were conducted. The one, denoted as *fit M*, is the result of ordinary χ^2 minimization (with the help of the Minuit package) on a set of resonances derived from previous studies [9, 12]. In the second one, denoted as *fit L*, LASSO regularization was applied, combined with the aforementioned information criteria, leading to a sparser model.

Apart from the coupling constants, cut-off values for hadron form factors are included as free parameters in our model. Resonances with spin-1/2 provide one coupling constant and resonances with spins-3/2 or 5/2 provide two coupling constants, each. Masses and widths of the resonances were taken from [13]. The results of the fitting procedures are summarized in Table 1.

Table 1. A summary of the features of the two fits.

	M: Minuit	L: LASSO + IC
no. of resonances	14	9
no. of parameters	25	17
χ^2 / n.d.f.	2.4	3.2

The photon beam asymmetry $\mathbf{\Sigma}$ is defined as

$$\mathbf{\Sigma} = \frac{d\sigma^{\perp} - d\sigma^{\parallel}}{d\sigma^{\perp} + d\sigma^{\parallel}} \tag{4}$$

where $d\sigma^{\parallel}$ denotes the differential cross section when the incident photon beam is linearly polarized in the x-direction and $d\sigma^{\perp}$ the differential cross section

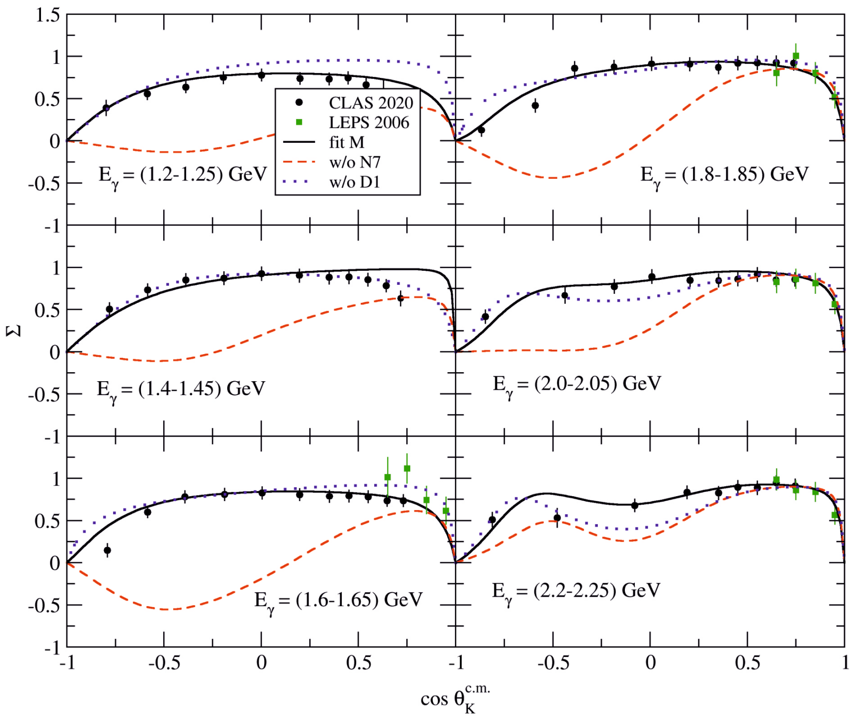

Figure 2. Photon beam asymmetry data compared to the full *fit M* results (solid line). Same fit, with the $N(1720)\ 3/2^+$ (dashed line) and $\Delta(1900)\ 1/2^-$ (dotted line) resonances omitted. The data are from CLAS [6] and LEPS [7] experiments.

when the photon beam is polarized in the y-direction. Figures 2 and 3 (taken from [5]) demonstrate how the results of the two fitting procedures M and L compare to the data on the Σ asymmetry. A useful observation that can be drawn from these figures is the importance of the $N(1720)\ 3/2^+$ resonance, given the disagreement with the data when it is omitted from the fits.

5 Conclusion

In our study of kaon photo-production off a neutron target with an isobar model we have incorporated recent polarization data and have employed a novel approach to fitting. The application of LASSO regularization in combination with information criteria provides a tool for automatic selection of parameters based on their information content. This turned out to be especially useful in our case, where the number of possible combinations of resonances is prohibitively large in the region of interest.

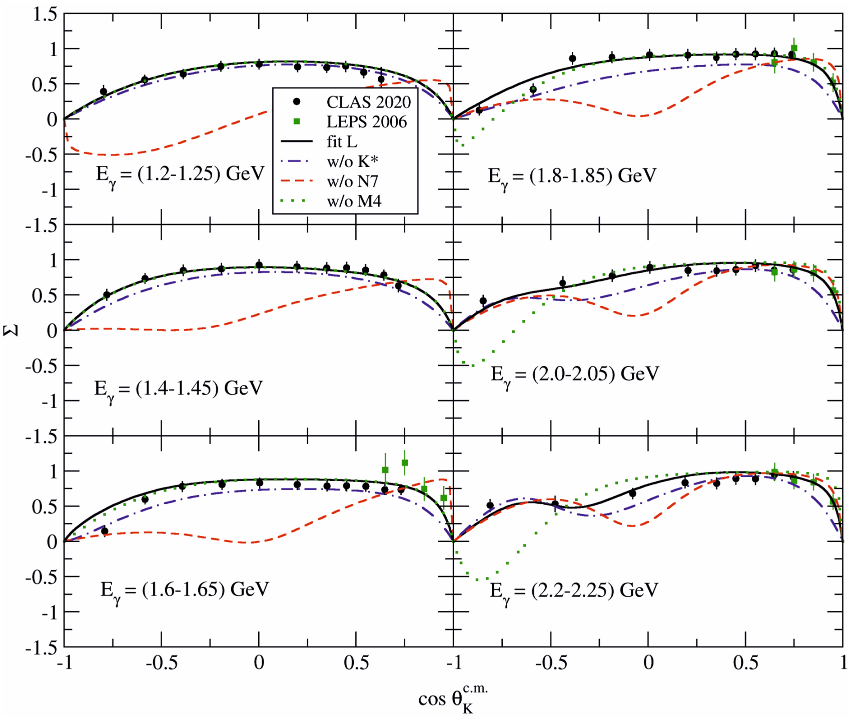

Figure 3. Photon beam asymmetry data as in Figure 2, compared to *fit L* results (solid line). Same fit, with the $K^*(892)$ (dash-dotted line), $N(1720)\ 3/2^+$ (dashed line) and $N(2060)\ 5/2^-$ (dotted line) resonances omitted.

Acknowledgements

D. Petrellis gratefully acknowledges support from the Bulgarian National Science Fund (BNSF) under Contract No. KP-06-N48/1. This work was supported by the Czech Science Foundation GACR Grant No. 19-19640S.

References

[1] B. Guegan, J. Hardin, J. Stevens, and M. Williams, *JINST* **10** (2015) P09002.

[2] J. Landay, M. Döring, C. Fernández-Ramírez, B. Hu and R. Molina, *Phys. Rev. C* **95** (2017) 015203.

[3] J. Landay, M. Mai, M. Döring, H. Haberzettl and K. Nakayama, *Phys. Rev. D* **99** (2019) 016001.

[4] T. Hastie, R. Tibshirani and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2^{nd} ed., Springer Series in Statistics, New York (2009) Chaps. 3, 7.

[5] P. Bydžovský, A. Cieplý, D. Petrellis, D. Skoupil and N. Zachariou, *Phys. Rev. C* **104** (2021) 065202.

[6] N. Zachariou *et al.*, *Phys. Lett. B* **827** (2022) 136985.
[7] H. Kohri *et al.*(LEPS Collaboration), *Phys. Rev. Lett.* **97** (2006) 082003.
[8] D. Skoupil and P. Bydžovský, *Phys. Rev. C* **97** (2018) 025202.
[9] D. Skoupil and P. Bydžovský, *Phys. Rev. C* **93** (2016) 025204.
[10] H. Akaike, *IEEE Transactions on Automatic Control* **19** No. 6 (1974) 716-723.
[11] G. Schwarz, *Ann. Stat.* **6** No. 2 (1978) 461-464.
[12] L. De Cruz "Bayesian model selection for electromagnetic kaon production in the Regge-plus-resonance framework," Ph.D. Thesis, Ghent University (2012).
[13] P.A. Zyla *et al.*, *Particle Data Group, Prog. Theor. Exp. Phys.* **2020** (2020) 083C01.